

BEST PRACTICES in BIG DATA STORAGE

PRODUCED BY:
Tabor Communications Custom Publishing Group

IN CONJUNCTION WITH:



Best Practices in Big Data Storage

Conducted by: Tabor Communications Custom Publishing Group

Dear Reader,

From Wall Street to Hollywood, drug discovery to homeland security, companies and organizations of all sizes and stripes are coming face to face with the challenges – and opportunities – afforded by Big Data. Before anyone can utilize these extraordinary data repositories, however, they must first harness and manage their data stores, and do so utilizing technologies that underscore affordability, security, and scalability.



In Best Practices in Big Data Storage, we review the relentless growth in data across a variety of vertical markets, discuss the issues presented by managing petabytes of data, and detail innovative technologies for big data storage from five key vendors.

Please enjoy reading this special report with expressed hope that some of what you might learn from it can be used in your own professional endeavors.

Cordially,

Kevin Davies PhD

About the Author:

Kevin Davies PhD is a science writer and author. He is the founding editor of Bio-IT World magazine, which covers the intersection of IT and life sciences, and a former editor at the prestigious scientific journal Nature and editorial director at Cell Press. He is also the author of three books, most recently "The \$1,000 Genome" (Free Press, 2010).

Study Sponsors



Media Partners





Best Practices in Big Data Storage

“From the dawn of civilization until 2003, humankind generated 5 exabytes of data. Now we produce 5 exabytes every two days... and the pace is accelerating.”

— Eric Schmidt, Executive Chairman, Google

By Kevin Davies PhD

Introduction

“In 2011,” says venture capitalist and author Juan Enriquez, citing an IDC report, “we played, swam, wallowed and drowned in 1.8 zettabytes of data... if you were inclined to store this data on 32-gigabyte iPads, you would need only 86 billion devices – just enough to erect a 90-foot-high wall 4,000 miles long from the bottom of your shoes to the center of the Earth. Today, a street fruit stall in Mumbai can access more information, maps, statistics, academic papers, price trends, futures markets, and data than a U.S. president could only a few decades ago.”

Whether the eye-catching pronouncement of Google’s Eric Schmidt about mankind’s Exabyte explosion of Big Data is entirely accurate, there is no doubt that the era of Big Data has arrived, transforming almost every industry one can think of – from oil and gas and the financial markets to medical research and movie making.

“There is no real way to characterize ‘big’ in big data, it depends on the organization,” says Chris Blessington, EMC Isilon director of marketing communications. “If you’re a life science research firm generating [multiple] petabytes of data through genetic sequencing, that’s big data. If you’re a financial institution looking to optimize trading practices from a hedge

fund or a Monte Carlo simulation... that’s a completely different thing. It’s equally important, though a much smaller amount of data.”

In other words, the definition of ‘big data’ is completely relative to the organization in question and what it is hoping to accomplish, whether it is monetization, building a competitive advantage, or pioneering new research. “Big data is a concept that is about deriving value from information,” says Blessington.

The tipping point came about five

“There is no real way to characterize ‘big’ in big data, it depends on the organization.”

years ago, when the data deluge evolved from being a problem to an opportunity. But organizations seeking to derive value from their data must first harness, store and manage their data in an efficient and cost-effective way, before they can begin to apply new analytics methods and extract.

We should point out that not everyone is a fan of Big Data. “With

big data, researchers have brought cherry-picking to an industrial level,” wrote Nassim Taleb, New York University professor and author of *Antifragile*. “Modernity provides too many variables, but too little data per variable. So the spurious relationships grow much, much faster than real information.”

In other words, says Taleb: “Big data may mean more information, but it also means more false information.”

That is undoubtedly true, but the task at hand is to coral the data tsunami to build new insights and business opportunities. In that, users and vendors are in a close alliance to optimize new technologies for data management and analytics.

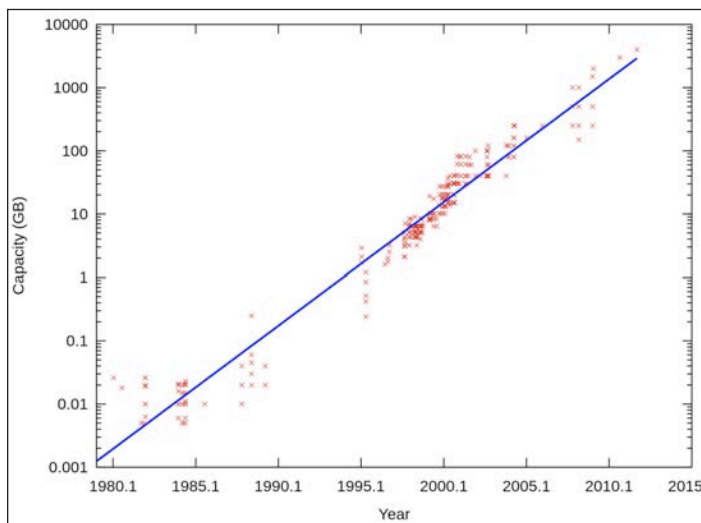
In this report, we will briefly review the scales of ‘big data’ before examining some of the technology solutions and trends that will help organizations marshal their data repositories for maximum impact and efficiency.

Brief History of Big Data

There is a photograph doing the rounds, taken in 1956, of a large container being loaded into an airplane’s cargo hold with a forklift truck. The initials “IBM” can just be made out on the top of the box — a state-of-the-art computer hard drive capable of holding 5 Megabytes. Today, that’s a disposable Justin Bieber track on your son or daughter’s iPod.

The astonishing growth in disk drive capacity, which has culminated in recent years with the worldwide deluge of data commonly referred to as Big Data, is encapsulated in Kryder’s Law, as suggested by Seagate engineer Mark Kryder. Modeled on Moore’s Law, Kryder’s Law points to the inexorable growth of hard drive capacity since the introduction of the disk drive in 1956, but particularly over the past 30 years.

FIGURE 1. Kryder’s Law Shows Exponential Increase of Hard Drive Capacity Over Time



Kryder’s Law, named after Mark Kryder, shows that hard drive capacity is growing over time faster than Moore’s Law. Source: Wikipedia

The Genome Generation

Ten years ago marked the official completion of the Human Genome Project (HGP), a \$3-billion international effort to read out the full DNA sequence of man’s genetic code. There was just one problem: completing that reference sequence required more than a decade and armies of scientists running industrial warehouses full of scores of outdated sequencing machines.

Even as President Clinton celebrated that historic landmark, scientists and entrepreneurs were busy conceiving new technologies to make good on the promise of the HGP and usher in a new era of personalized medicine. The “next generation sequencing” (NGS) technology had to be capable of decoding not one human genome per decade, but hundreds or thousands of genomes in a week (or preferably a day) for just a few thousand dollars.

The first NGS instrument debuted in 2006, but it was the emergence of a British technology developed by Solexa, now part of Illumina, that truly opened the floodgates. These instruments could generate many gigabytes of sequence data in a single run. Today, a single Illumina HiSeq 2500 can sequence two human genomes in a week, generating close to 1 Terabyte data in the process.

Almost overnight, any medical lab with an NGS instrument became the equivalent of a genome center. And the world’s leading genome centers – at the Broad Institute (Cambridge MA), Washington University (St Louis), Baylor College Medicine (Houston) and the Wellcome Trust Sanger Institute (UK) – became genome centers on steroids. Currently the world’s largest genome center is in Hong Kong, managed by BGI (the Beijing Genomics Institute).

Handling the prodigious data output of these instruments is a challenge for life scientists around the world. The leading genome centers currently house more than 10 PB data, but even smaller research institutes such as the Jackson Laboratory in Bar Harbor, Maine and the J. Craig Venter Institute in San Diego, are pushing into 1-2 PB territory.

The National Cancer Institute recently partnered with a leading systems integrator, Accunet, to build a new data center of at least 3 PB in a contract worth \$15 million. The data-center design included storage as well as switching, routing, wireless, virtualization, data optimization, and cloud connectivity. The new data center “will enable us to keep pace with the escalating amounts of biomedical data that our scientists work with every day,” said Greg Warth, the center’s IT operations director.

The major genome organizations have worked with numerous storage vendors. But IT budgets, particularly in the age of the sequester, are strained, putting a premium on data management. New technologies for tiering data and organizing metadata are crucial for effective management, as well as cultural “carrot-and-stick” incentives to curb scientists’ insatiable appetite for data.

The density of information a hard disk can record has grown about 50 million times, from a mere 2,000 bits to more than 100 billion bits (gigabits).

Another way of viewing this remarkable trend is in the

plummeting cost per unit of storage. In 1981, one gigabyte cost \$300,000. By 1990, that had fallen to \$10,000. In 2000, the cost/GB was \$10. Today, it's less than 10 cents.

MIT research scientist Andrew McAfee observes that the extraordinary growth in worldwide data over the past three decades not only shows no sign of abating, it may eventually require the invention of an entirely new terminology. In 1979, a firm called Teradata chose its name to reflect the first waves of big data. In 2008, *Wired* magazine coined the term "The Petabyte Age", only to be trumped four years later when its British sister publication *Wired UK* proclaimed "The Exabyte Revolution." That coincided with a report from Cisco heralding "The Zettabyte Era."

Zetta- is equivalent to 1 sextillion (10 to the power 21). The only other metric prefix is the Yotta- (septillion). A student at UC Davis proposed that the next level be termed "hella-". 1 hellabyte would be 1027 bytes. (No word as yet on whether that term has gained official recognition.)

TABLE 1. The scale of Big Data in Healthcare

Example	Bytes
Electronic Medical Record	1 Million Megabyte 1 Billion Gigabyte
Patient Genomic Data	1 Trillion Terabyte
A Genome Center	10 ¹⁵ Petabyte
Hospital (1M patients)	10 ¹⁸ Exabyte 10 ²¹ Zettabyte
Every person in the world sequenced!	10 ²⁴ Yottabyte

From Petabytes to Exabytes

Some organizations have already crossed the Exabyte threshold such as Facebook, which stores nearly a quarter of a trillion images. Yahoo, a customer of Data Direct Networks, is almost at that same threshold.

The traditional areas of Big Data are well known: financial services, oil and gas, media/entertainment, government/defense, R&D, healthcare/life sciences. But as highlighted in a recent Wall Street Journal article, Big Data is an increasingly common theme in more mundane settings, such as human resources, product development, and marketing.

The entire rendering of James Cameron's science fiction epic *Avatar* required more than 1 PB storage. The film was rendered in New Zealand using a computing core of 34 racks, 40,000 processors and 104 TB RAM. The blades were reading and writing against 3 PB fast fiber channel disk NAS from NetApp and others.

Few areas have felt the onslaught of the Big Data tsunami as much as the life sciences. Data generation exploded in 2006-07 with the arrival of next-generation sequencing,

which almost overnight pushed the amount of DNA sequence data being pushed out by labs worldwide by orders of magnitude (see Box).

More and more companies and organizations are now having to store and manage many petabytes of data, when they were comfortably dealing with mere terabytes just a few years previously. Here are just a few examples:

- UPS stores more than 16 PB data, from deliveries to event planning
- Monster, the online careers company, stores 5 PB data, largely from nearly 40 million resumes
- Zynga stores 3 PB data on the gaming habits of nearly 300 million monthly online game players
- Facebook adds 7 PB storage every month onto its exabyte trove
- The Boeing 787 Dreamliner generates 1 TB data for every roundtrip, equating to hundreds of TB daily for the entire fleet
- CERN has collected more than 100 PB data from high-energy physics experiments over the past two decades, but 75 PB comes from the Large Hadron Collider in just the past three years.

Not before time, there is now an academic *Journal of Big Data*. The new peer review, open access journal, published by Springer Verlag, examines the key challenges in big data, from data capture and storage to search, analytics, data mining and visualization, machine learning algorithms, cloud computing and distributed file systems.

Scale Out Storage

It is a testament to how quickly technology marches on that, according to Chris Dagdigan, a leading HPC consultant with the BioTeam, petabyte-capable storage is now relatively trivial to acquire. That said, the cost of ingesting data is falling faster than the rate at which industry is increasing drive capacity, which poses daunting challenges for users. "The [science] is changing faster than we can refresh datacenters and research IT infrastructure," says Dagdigan.

“The [science] is changing faster than we can refresh datacenters and research IT infrastructure.”

Those sentiments are echoed by the former director of research computing at Harvard University, James Cuff (now chief technology officer at Cycle Computing). In 2006, Harvard boasted a state-of-the-art, 30-TB local network-attached storage array. Today, the data center houses 25,000 processors and 10 PB storage. "And that's just a small, university-wide research computing offering," says Cuff.

Cuff says the component parts of storage and process-

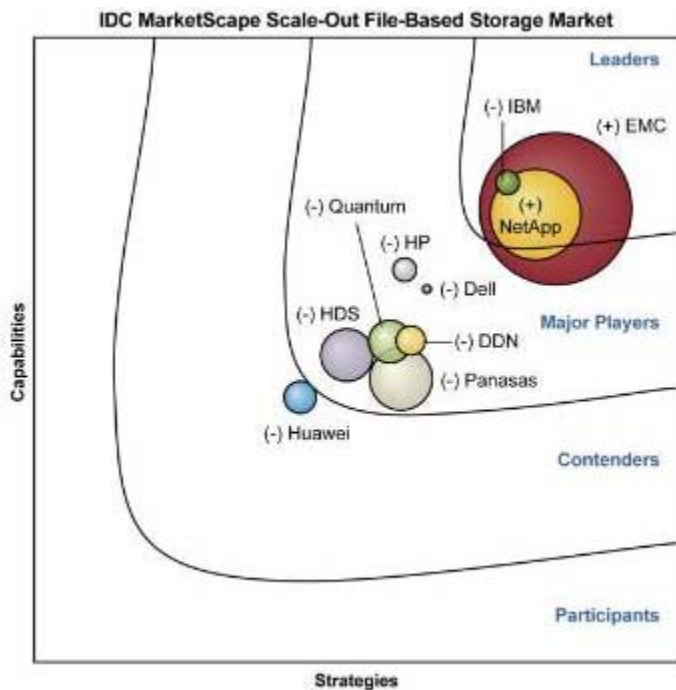
ing are largely solved problems at this point. “Everyone is collecting Big Data, but this collection phase is a prelude to a second phase—namely once collected, trying to work out what we ought to do with it so history informs the future. ‘Big data’ is a very hyped term, but it’s real.”

Over the past five years or more, dozens of vendors have offered scale-out, file-based storage solutions to handle the Big Data onslaught. The big names – Sun, NetApp, EMC etc – have been joined by a fleet of innovative storage providers, including DDN, Panasas, Quantum, Cray and others.

Some of these technologies have attracted massive sums in acquisition. In 2010, EMC acquired Seattle-based Isilon Systems for \$3 billion, while Hitachi Data Systems bought San Jose-based BlueArc, which features FPGAs (field-programmable gate arrays) in its systems.

The attributes of several of these leading companies – Cray, DDN, NetApp, Panasas, and Cleversafe – can be found elsewhere in this report. Company executives lay out their firm’s approach to key issues such as scalability, management, provenance and efficiency. (See BOX)

FIGURE 2. IDC Evaluation of Scale-Out Storage Vendors



Source: IDC MarketScape report (2012)

According to a 2012 IDC MarketScape report on worldwide scale-out storage:

“Incumbent storage suppliers with a proven track record in the industry are using scale-out solutions to expand market reach beyond the capabilities of their traditional scale-up offerings... On the other hand, suppliers with a core focus on data protection or high-performance computing are leveraging scale-out solutions to further

Vendor Views

This report highlights the groundbreaking technology solutions from five vendors working with a wide range of organizations and companies.

- **Cray** is widely known as a supercomputing powerhouse, but that is just one of the firm’s three divisions, the others focusing on storage and analytics. The company that supplies the muscle for the fastest supercomputer in the world right now (Titan in Tennessee) also sells Sonexion, a scale-out storage system for Lustre, and partners with other storage providers. “We’re much more of a holistic systems company,” says Cray’s Jason Goodman. “Customers value not just that our systems perform as advertised but we’re a trustworthy partner in their infrastructure business.”
- **NetApp** is one of the most respected and established names in big data storage, with two excellent lines of storage products to suit a range of user needs. A recent acquisition gave rise to the E-Series, which delivers a potent combination of performance, density and energy efficiency. The company is also a major advocate of flash storage, saving space and power consumption, including all-flash arrays. The company boasts scores of clients with more than 100 PB storage.
- Chris Gladwin, CEO and founder of Chicago-based **Cleversafe**, says he wants to establish his firm’s technology as the “de facto industry standard.” The company sold more than 100 PB storage in 2012 featuring its novel dispersed computing technology, which slices data and distributes it for more effective and secure reconstitution. A marquee project with Lockheed could set the stage for breakout sales across the industry. Not surprisingly, the firm has the backing of the city’s Mayor, Rahm Emanuel, who says “Cleversafe is the kind of business we think of when we talk about Chicago as the city of big data!”
- **DDN** is a rapidly growing storage company with a strong tradition in the HPC market. Unlike some ‘big iron’ manufacturers that rely on acquisitions for innovating, DDN is a company with “a history of real innovation and a strong market focus that’s funded by a profitable business,” according to a leading industry analyst. DDN is agnostic about many key aspects of storage infrastructure, including the file system and choice of interconnects.
- The **Panasas** ActiveStor parallel storage product line is winning admirers across continents and industries. The company says it provides “plug-and-play simplicity” for technical computing and big data workloads. Users tout the technology’s reliability and dependable storage performance. “ActiveStor is fantastically easy to configure and manage,” says the IT manager at a leading UK university. “We don’t have a dedicated storage administrator, so it’s essential that our systems don’t take a lot of time and effort to manage.”

penetrate their established use cases in their respective core customer base.”

IDC expects the market to grow for several more years, in part because “scale-out systems are becoming increas-

ingly object aware from a data organization and access perspective.”

The medium of data storage has improved dramatically in recent years. For an in-depth report on the technology of Hadoop, see the recent Tabor Communications report, [“Big Data Implementation: Hadoop and Beyond”](#) by Mike May.

Earth to Cloud

The monuments of Big Data can be found in the large data centers dotting the country, including some found in surprising locations.

In Manhattan, Sabey has just opened Intergate.Manhattan, a stunning data center of 32 floors near the Brooklyn Bridge. The building’s first tenant is the New York Genome Center, a state-of-the-art genome center formed by a consortium of 11 New York medical and research centers. When fully operational, the Sabey’s Intergate.Manhattan will accommodate 40 Megawatts of critical data center capacity on 600,000 square feet of data center floor space.

In downtown Boston, the Markley Group owns a major data center in the same building as a famous department store. Although the firm’s reputation is in colo facilities, it recently launched a Cloud Services initiative. “Data center customers who buy cage space are all asking for additional services — it’s ‘bring me a cloud or I’m leaving’,” says 451 Group analyst Carl Brooks.

The initial Markley Cloud implementation builds on VMware technology and features NetApp storage. Customers availing themselves of Markley’s facilities and resources in bandwidth, power and cooling include Oracle, Vertex Pharmaceuticals, Harvard University and the Broad Institute.

The IDC MarketPlace report mentioned earlier predicts that adoption of the cloud will be an attractive rout for many midsize organizations. Of course, one of the transcendent trends in data storage has been the cloud, chiefly that of Amazon Web Services and its Simple Storage Service (Amazon S3). Amazon S3 stores more than 900 billion objects, at a total storage size of about 100 PB.

(Many experts have touted Amazon’s unparalleled infrastructure and security, but even Amazon is not infallible. Two years ago, power outages temporarily shut down several social media websites. And in a recent S3 security breach, thousands of buckets – and more than 100 billion files — were unwittingly left open to the public when the relevant storage accounts were not set to private. Files at risk included car dealer sales, employee spreadsheets, unencrypted and a videogame producer’s source code.)

At the other extreme, advances are being made in archived storage and backup. In 2012, Amazon introduced Glacier, a very low-cost, cold storage service designed for archiving data and backup purposes in cases where users seldom need to access the data. Storage costs run as low as 1 cent per gigabyte per month, providing a dramatic cost savings compared to traditional datacenter tape storage.

The advantages of Glacier are many. It spares organizations from the upfront capital expenditure and the recurring oper-

ational costs such as power, personnel and maintenance. It also avoids the common problem of over-provisioning – acquiring more storage at greater cost to ensure the organization has sufficient capacity on hand. Amazon says Glacier “changes the game for data archiving and backup” as users pay nothing upfront, a minimal price for storage, and can scale their capacity requirements as needed.

That said, Glacier isn’t the Big Data archiving solution for everybody. It takes several hours to retrieve files, which didn’t suit the needs of Facebook as it sought to store 1 exabyte of rarely accessed digital photos — some 240 billion images (350 million added daily). The solution was a dedicated “cold storage” facility in Prineville, Oregon, with no generators but 1 exabyte capacity. The facility holds 500 racks, each containing 2 PB data. Each rack consumes just 2 kW power (compared to 8 kW in a standard rack). The hardware uses shingled magnetic recording to fit more data tracks.

Facebook’s challenges are orders of magnitude beyond most organization’s experience, but as Jay Parikh, Facebook’s VP Infrastructure Engineering, says ominously:

“Our big data challenges that we face today will be your big data challenges tomorrow.”

Data Management

“The problem of big data is not only that it is capacious, but that it is also heterogeneous, dirty, and growing even faster than the improvement in disk capacity,” says David Patterson, computer scientist and professor at UC Berkeley. A major challenge then is “to derive value by answering ad hoc questions in a timely fashion that justifies the preservation of big data.”

“ The problem of big data is not only that it is capacious, but that it is also heterogeneous, dirty, and growing even faster than the improvement in disk capacity. ”

Patterson has established a new group at Berkeley called the AMPLab, which combines statistical machine learning (Algorithms), cloud computing (Machines), and crowd-sourcing (People). An early priority is cancer genomics, because Patterson feels that the rigor of expert software and computer engineering is needed to complement recent progress in genome sequencing.

One idea, proposed by David Haussler (UC Santa Cruz) is to build and then mine a data repository of 1 million cancer genomes in their entirety (see Table). Another new project, launched by the American Society for Clinical Oncology is called CancerLinQ. Researchers will mine medi-

cal data of more than 100,000 cancer patients from medical records.

TABLE 2: Cancer Genomics Hub (CG Hub)

- Repository at SDSC to store Cancer Genome Atlas data
- 50,000 genomes (100 Gigabytes/genome) = 5 Petabytes
- Annual cost = \$100/genome or \$5 million (50,000 genomes)
- Co-location opportunities in same data center for groups who want to compute on data

Credit: D. Patterson (AWS 2012)

As the volumes of data repositories swell into tens and hundreds of Petabytes – and millions of individual files – more and more organizations are confronting near insurmountable challenges in locating and retrieving specific files. Although storage vendors are providing greater accessibility to data, a growing number of organizations are turning to dedicated data management software solutions.

A key to success in data management comes down to policies and culture that demand users vigilantly monitor their individual data repositories and discard data that are no longer required. This may involve some sort of “charge back” scheme, as deployed by the Broad Institute for example, which directly engages researchers “in the process of claiming ownership or disallowing interest in data files,” says Trunnell. In 2012, the Broad was actually able to permanently delete a small portion of its data – a fairly traumatic undertaking for any scientist!

A growing trend is in tiering. “We’re back to needing storage tiers. I need high performance *and* storage,” says BioTeam’s Dagdigian, who is making greater use of Tier 2 storage, shifting to multiple vendors and mixing high- and low-end products as necessary.

A further trend in data management is that the latest storage offerings from DDN, Panasas, and others can run Unix on standard architectures. In other words, “your storage will be able to run apps,” says Dagdigian.

But increasingly users are going to need dedicated software to manage their petabyte repositories. For several years, General Atomics (GA) has partnered with the U.S. Department of Defense (DoD) with its Nirvana SRB (Storage Resource Broker) software. The DoD has found Nirvana SRB to be an excellent solution to manage and organize files, which are currently around the 40-50 PB range, in areas such as intelligence gathering and battlefield simulation.

Tools such as Nirvana SRB allow users to tag and federate their data, improving productivity and adding value to the data. It also provides hierarchical storage management (HSM) and the ability to easily locate and pull up archive data.

But some Big Data companies tend to assume that phar-

maceutical companies are stacked with software developers who understand Hadoop – and that’s just not the case, says Dagdigian. “There are maybe one or two people who understand Hadoop in the entire company, so selling a six-figure toolkit is not going to go over very well.”

For users who can’t afford the steep licensing fee, an alternative is open-source software called iRODS (Integrated Rule-Oriented Data System). The middleware software prepares an inventory of a user’s files by crawling through the file system and making a database entry for each file and directory.

“iRODS is very good,” says Chris Smith, co-founder of Distributed Bio, a San Francisco consulting firm. “The systems using it now are such large scale, with multiple sites, that it’s the only kind of system that really works.”

Smith, a veteran of Platform Computing, recalls seeing customers struggling to manage data distributed across multiple file heads or a large cluster. “I thought of my digital camera problem,” he says. “That layer of manageability was missing. iRODS is a key piece that helps manageability.”

iRODS has been widely tested in organizations from genome centers such as the Wellcome Trust Sanger Institute in the UK to large national and multi-national projects, such as the French National Library, the Southern California Earthquake Center, the UK e-Science program, and the High Energy Accelerator Research Organization in Japan.

For many organizations, however, an open-source tool is simply not feasible, and needing additional personnel to install, maintain and support the product. More commercially-accessible versions are under development from third parties such as Cambridge Computer.

Geek Squad

There is a sense among some Big Data leaders that the infrastructure challenge has largely been met. But as the volumes of data swell into exabyte territory for more and more organizations, the biggest challenge is going to be devising ways to mine the data and make sense of it all, or at least in part – to turn data into knowledge, and knowledge into wisdom.

“ There are maybe one or two people who understand Hadoop in the entire company, so selling a six-figure toolkit is not going to go over very well. ”

We “are relatively naïve in the way we represent and interact with the data,” says Matthew Trunnell, CIO of the Broad Institute. Because the research field is moving so

fast, “We don’t have the luxury, or have not taken the effort, to optimize algorithms for infrastructure.”

It is a business opportunity as big as the data stores themselves. As the *New York Times* pointed out recently, “Silicon Valley is full of Big Data start-ups that promise to translate the ever-growing flood of digital data into practical insights in business and science.” So what differentiates some of these firms, with quirky, enigmatic names such as Ayasdi, Cloudera, Palantir, MarkLogic and Splunk, which are providing exciting and novel ways to analyze Big Data?

Ayasdi, which means “to seek” in Cherokee, is a 2008 spin-off from Stanford University, applies research from DARPA and NSF to solve complex problems. Co-founded by mathematics professor Gunnar Carlsson, the firm’s technology is based on a technique called topological data analysis, which analyzes the ‘shape’ of complex datasets to identify trends and clusters that are statistically significant, without the user necessarily knowing which questions (s)he should be asking.

“The idea is to get the data to speak to you in a more unfiltered way,” says Carlsson.

“There’s a big bottleneck in a lot of different science labs, from astronomy to high-energy physics to oceanography.”

Adam Ferguson, a neuroscientist at the University of California, San Francisco, has used Ayasdi’s software to query large biomedical databases and is encouraged at the ability to detect and present patterns. He says the software is qualitatively different than other commercial data analysis tools or R, the open-source programming language. Early customers include Merck and the Food and Drug Administration in areas from drug discovery to oil and gas and finance.

Cloudera – the brainchild of former Facebook scientist Jeff Hammerbacher – saw a need for new open-source tools to help do science faster. “There’s a big bottleneck in a lot of different science labs, from astronomy to high-energy physics to oceanography,” Hammerbacher told PBS’ Charlie Rose. “They’re generating tremendous volumes of data and didn’t have the software and hardware infrastructure to be able to capture that data effectively.”

In partnership with Eric Schadt at Mount Sinai School of Medicine in New York, Cloudera is building a scalable infrastructure for data storage and analysis, which ultimately could aid healthcare delivery, the discovery of new diagnostics and therapeutics, and integrate molecular and clinical data. Ultimately, Hammerbacher says he would love to study the brain – how it works and how it breaks.

Palantir, which also has Stanford roots, was founded a few years earlier, when billionaire Peter Thiel (Founder Partners), a co-founder of PayPal and an early investor in Facebook, pitched an idea to his former college friend Alex Karp: Could they build software that deployed the same sort of approach PayPal had used to combat financial fraud to identify and thwart terrorist networks?

Palantir’s name comes from the “seeing stones” in the *Lord of the Rings*. The firm’s software uses a tagging technique to log every scrap of data separately and connecting different databases, and is proving particularly useful for the intelligence community, where a plethora of databases and security clearances hinders the aggregation of data and the identification of red flags that could prevent a terrorist attack.

The company has quickly found a niche in the government and security arena, working for the CIA, FBI and the Pentagon, as well as a variety of foreign governments. It is also branching out into the financial industry and biomedical research.

Another promising start-up in Splunk, which collects, organizes and harnesses machine-generated Big Data from a variety of sources – websites, applications, servers, mobile devices, and so on – to enable organizations to search, analyze, and act on vast streams of real-time and/or archived data.

The software is already being used by more than 5,000 enterprises, universities, and other clients in nearly 100 countries. The company has been profitable since 2009 and conducted a successful public offering in 2012.

Conclusion

We have tried to emphasize in this report that the unremitting growth of Big Data is both a challenge and an opportunity. Many technology providers are working tirelessly to keep up with the demand for data storage, and to a large extent are succeeding. What remains to be seen is how a new generation of algorithms and data analytic strategies will be able to successfully mine Big Data repositories, potentially helping users find meaning without knowing what questions to ask.

As storage technology continues to evolve, scientists may end up turning to a surprising old medium for storing data – nature’s very own hard drive, DNA. On a recent television appearance, Harvard scientist George Church handed Stephen Colbert a slip of paper containing 20 million copies of his new book *Regenesi*s, digitally encoded in DNA. (Colbert was momentarily speechless.) British scientists similarly encoded a Shakespeare sonnet and a clip of Martin Luther King’s “I have a dream” speech.

The cost of DNA storage is still a little prohibitive – about \$12,000/Megabyte. For now. ■

To request a complimentary subscription to *HPCwire*, *HPC in the Cloud*, *Datanami* or other Tabor Communications e-newsletters go to... <http://www.hpcwire.com/xs/register>



Changing the way the world saves data.

When one company executive proclaims that it is poised to become “the Google and Apple of the storage industry,” and another says it is “changing the way the world stores data,” it is probably time to pay attention.

When MIT-trained engineer and serial entrepreneur Chris Gladwin sold his company MusicNow to Circuit City more than a decade ago, he wasn’t looking to jump into the storage business. But as he struggled to effectively and securely digitize his sizeable personal collection of music files and other media, Gladwin saw an opportunity to reinvent storage software.

“He saw a huge business opportunity,” says Cleversafe’s Russ Kennedy, VP, Product Strategy, Marketing and Customer Solutions. “Traditional systems to protect data were created 30 years ago, long before people knew what data would look like today.”

Founded in Chicago in 2004, Cleversafe honed its technology for a few years before launching its software commercially. The company now has 100 employees and is growing fast. Among its major customers and partners are the federal government, Shutterfly, and Gladwin’s former company Lockheed.

Cleversafe’s technology relies on dispersed storage, also known as information dispersal. “We’ve developed a system to pull down barriers and create a limitless storage system that can grow and grow,” says Kennedy. The company tagline is limitless data storage. “Our approach really does remove these limitations,” says Kennedy. “Now our customers can worry about more important things for their business.”

With data stores ballooning into multi-petabyte ranges, and with individual disks growing into several terabytes, the time to rebuild lost data is becoming a problem. Using traditional RAID storage technology, data distributed across replica drives can be rebuilt in the event of a partial disk failure. But if another drive fails while a rebuild is taking place, there is a good chance that potentially critical data will be lost.

“Companies are starting to replicate data,

which is no big deal if we’re talking a few hundred Terabytes,” says Kennedy. “But once you move into Petabytes or even Exabytes, the costs in power and pooling and management when you start copying that just breaks the bank.”

The patented Cleversafe technology relies on object storage. The company likes to use an analogy of valet parking, in which the object (the car in this case) is placed where it makes most sense for safety and easy retrieval.

For Cleversafe, the secret to scalable and secure storage lies in object storage, rather than traditional file or block storage. It’s an entirely new approach, says Kennedy, grounded in established networking and wireless communication workflows. The data are spliced up and dispersed in a single data center or multiple centers, geographically or in the cloud.

With each slice of data, a little extra data are created, so only a subset of the original data need to be recovered and reassembled. You “can have multiple failures, drives can fail, networks can go down, datacenters can be brought down, but as long as you have that minimum threshold to put the data back together, you can recreate your data perfectly,” Kennedy says.

Exactly how Cleversafe slices the data depends on a host of factors. A sample configuration uses the “10 of 16” maxim – when 16 data slices are spread across a network of storage nodes, Cleversafe only requires ten to reconstitute the data.

Game Change

Deploying Cleversafe technology enables organizations to tolerate failure. “It is much more cost effective, it provides much more reliability and scales infinitely. We are changing the way the world stores data,” says Kennedy.

Shutterfly – the online photograph storage and printing company – is one of Cleversafe’s marquee customers. “They make money when you take your uploaded photos and build a mouse pad or an apron for mom,” says Kennedy. “Today everyone’s got a camera on their phone. Their storage costs were growing exponentially. We know how to protect data in that realm.”

While Cleversafe’s largest clients typically

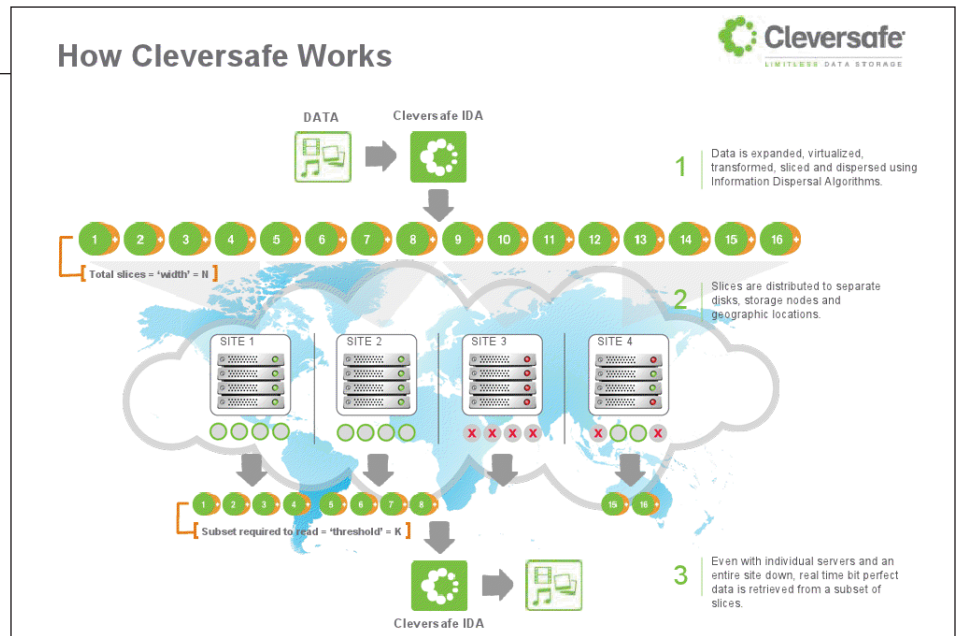
prefer to install the software on their own commodity hardware, many customers buy it in the form of an appliance. The Cleversafe hardware features Intel processors and standard disks. "Because it can tolerate failure, we can go with relatively low-cost commodity hardware pieces. It's our software that delivers the reliability," says Kennedy.

In addition, Cleversafe is an infrastructure provider to cloud providers, helping firms stay competitive with Amazon, improving their performance and reliability. "Many service providers want to create clouds. That infrastructure needs to be cost effective — and our technology allows that," he says.

Cleversafe is gaining traction in several key verticals, notably media and communications. Other booming areas include healthcare — think of the millions of patient images and scans that must be kept for the patient's lifetime — and the oil/gas and geospatial/satellite surveillance sectors. The company's main focus has been in North America, but it is seeing strong growth in the UK and Japan.

For the Federal Government, a tool like Cleversafe enables long-term data preservation for the intelligence community, which as numerous organizations, faces tough decisions about which data to keep and which to delete. With Cleversafe, organizations can keep all their data and analyze in totality over time. "It allows you to tease out a lot of intelligence because you can keep it all," says Kennedy. "The more data you can save, the better your analytics are, and the better your decision making is going to be."

Indeed, Cleversafe makes the case



that its technology actually delivers greater benefit and customer value as the data store gets bigger — in contrast to most other systems. As data stores grow, there's typically some sort of compromise — a "gotcha" as Kennedy puts it — in which organizations have either to pay more for increasing reliability and scalability, or make tradeoffs in the area of security. "We get more cost effective the bigger it gets, the more secure and reliable it gets," says Kennedy.

Cleversafe makes the eye-catching claim that its system is 100 million times more reliable than RAID. "It would be 100 million years before you would have a chance to lose data despite drive, server, network or site failures," says Kennedy. "If that means something to you — which for many industries it does — then our system is a great solution."

Cleversafe's software runs as a seamless system with no single point of synchronization or potential failure. It has three major components: 1) Out of band dsNet Manager; 2) Acceser, which does all slicing, dispersing and reading of the data; and 3) Slicestor — the storage nodes where the data sits. The system continually performs integrity checks in the background to sense if data have been compromised and then rebuilds from other slices, rather than waiting until a request is generated.

Big Data

A major partner of Cleversafe's is Lockheed. Often

datasets are too large to move into an analytical engine, but Hadoop is changing that equation. However, as Hadoop's architecture relies on replication, this requires three-fold redundancy to ensure availability. Cleversafe has replaced that underlying HDFS architecture that requires just one version of the data. This translates to a 3x increase in efficiency, as users can perform analytics on just one instance of the data.

"The analytics are now much better and deeper," says Kennedy. "We're enabling the analytics to be more reliable, deep and accurate." Cleversafe is also collaborating with Lockheed in building a cloud computing resource for the US Government.

No doubt many big storage companies are recognizing that dispersed storage technology is the wave of the future. That will either mean more opportunities for collaboration or more competition for Cleversafe.

"There are going to be trade-offs," says Kennedy. "Reliability and security will cost you or a lower cost system might lack reliability. If those trade-offs are fine for you, then Cleversafe probably isn't the system for you. If finding those trade-offs are pain-points, that's when you need us."

Further Reading: White Paper: "Cleversafe Object-based Dispersed Storage" <http://www.cleversafe.com/images/pdf/Silverton-Cleversafe-Object-Based-Dispersed-Storage.pdf> ■

www.cleversafe.com

Seamless Software

Cleversafe's software runs as a seamless system with no single point of synchronization or potential failure. It has three major components:

- 1) Out of band dsNet Manager;
- 2) Acceser, which does all slicing, dispersing and reading of the data; and
- 3) Slicestor — the storage nodes where the data sits. The system continually performs integrity checks in the background to sense if data have been compromised and then rebuilds from other slices, rather than waiting until a request is generated.



Applying the ABCs of Big Data to practical storage solutions.

Since its founding in 1992, NetApp (formerly known as Network Appliance) has remained at the innovative edge of storage technology, and was the only storage vendor on the 2011 *Forbes* list of the “World’s Most Innovative Companies”. Globally, the Fortune 500 company has more than 12,000 employees and 150 offices and reported more than \$6 billion in revenue.

NetApp is not new to the idea of Big Data either.

While other vendors might consider selling a few petabytes of storage a major success, NetApp’s biggest customer has more than 1 Exabyte of managed storage and more than 100 NetApp customers have 10 Petabytes (PB) or more.

NetApp pioneered the concept of storage as an appliance — specifically network-attached storage (NAS) — but is known for many other innovations over the years, including the WAFL (Write Anywhere File Layout) system for data protection, thin provisioning, and deduplication. Now it is championing practical solutions to Big Data.

“Our strategy with Big Data is no different than with our many other segments, which is to offer the best-of-breed data management and storage that supports an open ecosystem of partners to provide our customers complete solutions. Our deep integration with partners creates unique customer value and a preference for NetApp,” says Richard Treadway, NetApp’s Director of Big Data Solutions Marketing.

NetApp is the leading storage provider for the US Government. It supplies more than 50 PB storage for Sequoia, the former #1 supercomputer at the Lawrence Livermore National Laboratory. Other major verticals include healthcare, energy (20 of the leading US oil and gas companies), media and entertainment and a large majority of American financial institutions.

With such an array of clients, NetApp not surprisingly offers a range of platforms to manage different types of workloads — in-

cluding Big Data.

“We’re unique — we have a solution no matter what your application need is,” says LaNet Merrill, NetApp’s HPC Marketing Manager. “We can scale from very small configurations to very large in a modular fashion.” Merrill says NetApp offers the most efficient capacity footprint per dollar in the industry and world-wide support that smaller companies simply cannot match.

Contemporary storage means not just handling big data but also helping organizations manage data efficiently. “Our platforms are wicked fast,” says Merrill. It’s all about ingesting data and handling the data sharing and analysis, understanding parallel file systems and providing solutions.

NetApp offers two major platforms: Data ONTAP and E-Series. Data ONTAP is designed for “seamless scaling, continuous operations, and automated management.” This unified scale-out infrastructure solution allows users to deploy not only NAS and file-based protocols, but also SAN protocols such as Fiber Channel and iSCSI in a shared storage system.

The E-Series platform was introduced in with the 2011 acquisition of Engenio external storage systems from LSI. E-Series delivers extremely dense, high-performance storage. “We take advantage of the unique capabilities of both platforms when designing our portfolio of solutions for Big Data,” says Treadway.

ABC: Analytics, Bandwidth and Content

The term “Big Data” is confusing, says Treadway. In fact, he says, “it’s been called the most confusing term in IT taking over the mantle of #1 from ‘cloud’.”

Choosing the right storage infrastructure brings benefits to the top and bottom lines, from better business outcomes and revenue growth to improving storage efficiency. Treadway says NetApp defines a portfolio of Big Data solutions that address its divergent workloads. “We segment the portfolio into three areas we call the

ABC s of Big Data – Analytics (on large unstructured data with technologies like Hadoop), Bandwidth (for high performance dedicated workloads) and Content (for large content repositories and active archives).”

“With Big Data we see our customers often reach an inflection point where they can no longer grow incrementally,” he continues. It’s a point where they have to fundamentally think about doing things differently – as data sets grow to petabyte size that some tools to manage and protect them no longer work. It’s at this inflection point we help customers turn data growth into an asset not a liability.”

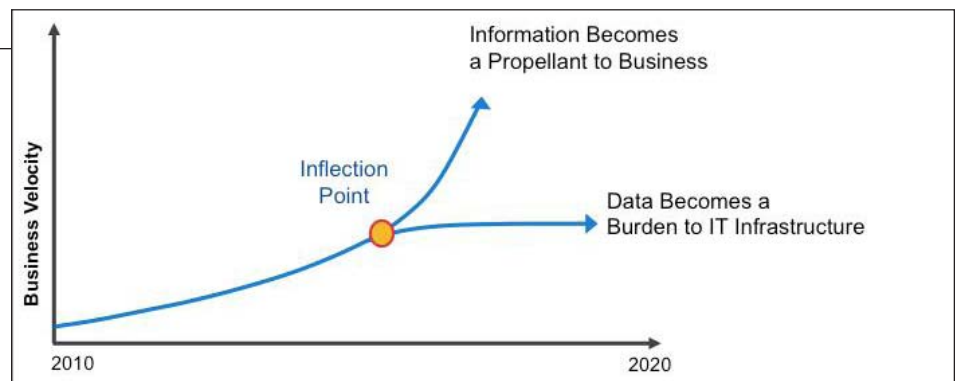
“There are three axes on which we see this inflection point occurring,” says Treadway. First is the increasing complexity of the data – images, voice and video. Second is the speed of data generation and processing. And third, the sheer volume of data, which must be stored virtually forever. As Treadway puts it, “The answer to Big Data can no longer be the delete key,” says Treadway. “Today’s data sets require a ‘keep forever’ strategy. For example a patient medical record must be kept for the life of a patient plus 20 years”

When NetApp engages with potential customers, the conversation is typically around “helping them get control of their data growth so that they can take advantage of the insight they find and turn them into action,” says Treadway.

A key aspect of leveraging Big Data is the ability to ingest at high speeds and use analytic tools to gain competitive advantage. NetApp strives to bring performance-driven data to enterprise users as well as HPC clients. “HPC is merging with Hadoop-style analytics,” says Merrill.

That may not impress HPC users so much, but more and more organizations are looking to use Hadoop to transform data and find the nuggets of insight that can help their business.

To ensure data reliability, the Hadoop best practice is to keep three copies of the data. But in the event of a failure, much of the processing is taken up with the recovery and that can lead to unpredictable job completion times. In some cases a single node failure can cause job completion times to more than double.



NetApp provides a Hadoop solution that uses E-Series storage arrays directly connected to compute nodes via SAS in a “shared nothing” configuration. Because the E-Series supports RAID 5 disc failure recovery is isolated to the storage array and as such Hadoop does not execute its recovery process. This results in job completion times under disc failure that are virtually identical to healthy cluster completion times.

This “Enterprise Class Hadoop” is critical when failure to deliver on SLAs results in substantial impact. In the financial sector, for example, being late on a hedge fund trade could result in losses of millions of dollars.

NetApp cites a case study closer to home: the company has a ‘phone home’ capability for all of its controllers – weekly reports spanning 20 years covering health and condition of each storage controller. Running reports on that amount of data (some 24 billion records) took 4 weeks in a traditional data warehouse environment, but with a 10-node Hadoop cluster, the same report took only 10 hours. An even bigger report (240 billion records) had not been possible, until it was run on Hadoop, when it ran in 18 hours.

What happens next?

Looking ahead, NetApp will continue to build out the Big Data Ecosystem featuring key strategic partners such as Oracle, Microsoft and SAP as well as smaller startups like Cloudera and Hortonworks. NetApp managers insist that vendors offering a complete integrated stack limits choice, which in a disrupted market like Big Data can be a real advantage. The downside of choice is the risk that the solution components won’t work together, but NetApp offers choice without risk through validation with its open ecosystem of technology partners.

Building that ecosystem around open standards means providing solutions including software, servers and networks. As an example NetApp collaborates with Cisco to deliver an integrated infrastructure of servers, network and storage called FlexPod.

“The applications that require Big Data are not entirely new,” says Dale Wickizer, NetApp’s Chief Technology Officer. “In the government realm, we have long dealt with the issues around significant demands on computing and content at peak moments—driven by weather events, elections and other political events, intelligence, surveillance and reconnaissance operations and space exploration, for instance. The infinite scalability of our storage, combined with the flexible approach of our partners, allows for a flexible infrastructure that can handle these demands without incident.”

Another trend is the use of Flash storage, something that has been in development since 2009. “At a certain capacity, SSDs [solid state drives] greatly improve the performance of the file system,” says Merrill. Outside HPC, there are transactional and latency-sensitive workloads that Flash helps performance.”

In short, NetApp is supplying proven solutions for handling Big Data and will continue to do so long after the term has disappeared and been replaced by next new market hype.

“I think in 4-5 years, we won’t be talking about Big Data anymore,” says Treadway. “Big Data is a way of doing business. When everyone is doing business that way, we probably won’t need a name for it!” ■

Further Reading: “World’s Most Innovative Companies” Forbes July 2011. <http://www.forbes.com/special-features/innovativecompanies>

www.netapp.com



Accelerating Mixed Workloads with Intelligent SATA/SSD Storage

With the launch of ActiveStor 14 at the Supercomputing conference in November 2012 (SC12), Panasas did more than simply provide an incremental improvement in performance and more cost-effective storage. By integrating flash-based solid state drives (SSD), its flagship ActiveStor 14 storage solution, provides high performance storage for small and large files alike.

This new fifth generation storage blade architecture offers powerful and flexible technology for a broad range of applications, from finance and life sciences to oil and gas and scientific simulations. It is the first product in the industry to use flash memory in an intelligent hybrid layout for mixed data workloads.

For many years, Panasas focused on data storage for high-performance throughput applications and put a premium on ease of use and manageability. The company's traditional verticals, which have expanded over the company's ten-year history, include high-performance computing (HPC) applications in energy, academia, and government/defense.

But of late, the company is increasingly focusing on the commercial HPC market, where it reports wins in manufacturing, finance/risk analysis, and life sciences, especially genomics. Panasas boasts strong sales in Europe and Asia as well as North America.

What makes the latest product in the ActiveStor lineup particularly notable is the way in which Panasas combines large-capacity Serial ATA (SATA) with SSD in an intelligent, unified fashion for a true hybrid solution.

"HPC storage is not just small or large files, it contains a mixture of both," says Panasas senior product marketing director, Geoffrey Noer. In many fields – life sciences being one – a high percentage of the files (say 70%) are very small, yet they take up less than 1% of the total storage capacity.

"If flash did not command a huge premium, all flash would be an excellent application," says Noer. Maybe one day, but for now, particularly in HPC environments, he says "you have to be very strategic in how you use

flash." In this way, with its hybrid solution, Panasas has successfully kept the dollar/terabyte equation at a viable level, indeed lower than in previous products, while incorporating sufficient SSD for HPC applications.

Panasas ActiveStor solutions employ object RAID storage. The beauty of that architecture is that the system controls where the files get stored. In this way, the smaller files can be stored on SSD while hard disk drives are better suited for larger files. While small files don't consume much storage capacity, they significantly impact performance, especially when executing tasks that utilize all the system files, such as replicating a volume to another storage system. And that is a perfect recipe for the use of flash.

Noer summarizes the strategy as follows: "If it is file-system metadata driving the namespace, or a small file less than 64K, put it on flash. If it is large file data, put it on the SATA disk. We've used SSD in a way that makes a really big difference in the responsiveness of the system without the cost penalty of an all SSD caching layer."

ActiveStor 14 delivers up to 1.4 million IOPS for 4K reads, while bandwidth scales to 150GB/second. Capacity scales to 8 petabytes of storage in a single global namespace. The net result is that ActiveStor 14 is very well suited for both large file throughput and small file IOPS. In short, it is excellent for HPC applications of all kinds. Indeed, some Panasas customers have scaled out to more than 100 shelves (up to 83TB per shelf).

Put another way, the improved performance is about "uncongesting" the small files and metadata by placing it on a separate tier, which is where the biggest performance enhancements are to be found.

Flash storage also enables faster system rebuilds. For example, ActiveStor 14 can rebuild 4TB drives in about the same time as the previous generation took to rebuild 3TB drives. Rebuilds are now typically measured in hours rather than the better part of a day

or two as is the case with traditional hardware RAID.

Parallel File Systems

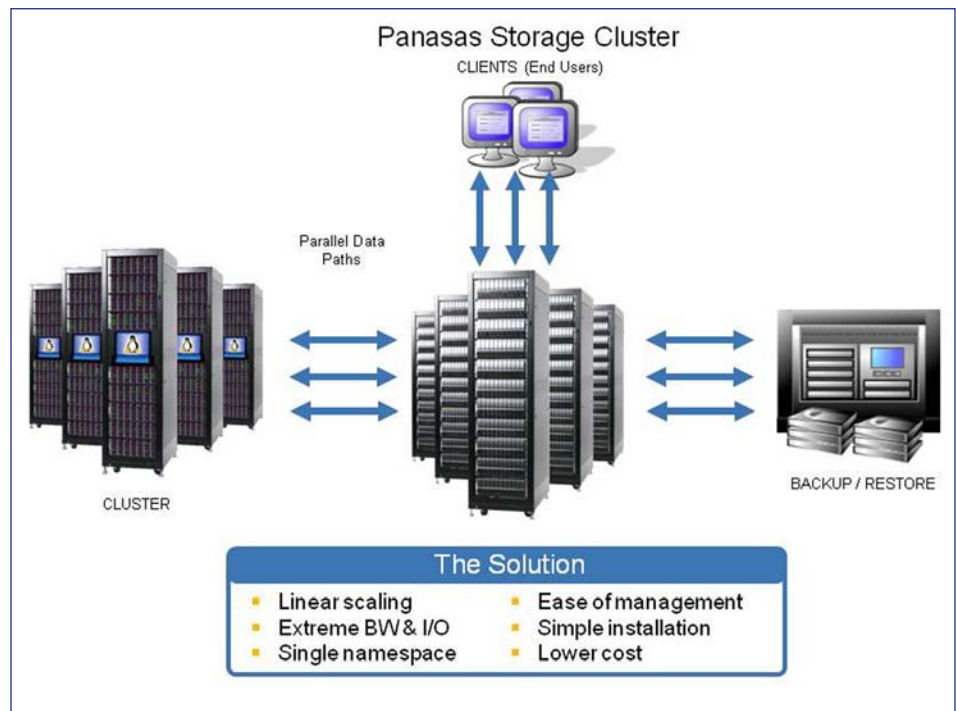
The Panasas parallel storage operating system, PanFS, shares a common heritage with the Hadoop File system (HDFS) – both were modeled on core research conducted by Panasas co-founder Garth Gibson at Carnegie Mellon University.

Because of the parallelism inherent in the two file systems, Panasas has been able to seamlessly integrate the ActiveStor infrastructure into a Hadoop environment. As a result, there is no need for users to buy dedicated compute or storage systems for Hadoop workloads— users can effortlessly run Hadoop on an existing compute cluster with Panasas storage.

Barbara Murphy, chief marketing officer at Panasas, says the company views the big data market in a couple of key ways: One is handling intensive workloads for deep analytics, such as running Hadoop on unstructured data workloads. Another is in the area of design engineering, aeronautics, , and scientific workloads such as genomics or simulation.

Most Panasas competitors are still operating on a legacy serial architecture that is some 20 years old, involving NFS access to a NAS filer, which in turn communicates to a hardware-based RAID controller and disks. Panasas, by contrast, was funded through government investment to build a parallel file system for massively parallel processing. Indeed, PanFS is one of three principle parallel file systems in use today; the others being open source Lustre and IBM's GPFS.

“Our approach was different,” says Murphy. “We combined the power of parallel processing with the ease of use found in a NAS appliance.” Compared to Lustre, which she says is a very complex system requiring expert personnel to manage, “you can run our system with very basic administration. We’ve taken the best of that core government research in parallel file systems, but brought it to the commercial space in a very simple appliance.”



The bigger the Linux cluster, the better, says Murphy – from 50 up to – thousands of cores suits Panasas very well. Because of the parallel architecture, the system can write straight to disk at very high speeds. “We get very high bandwidth, and now with ActiveStor 14, we can manage very high IOPS too,” says Murphy. “Moving to Panasas from other storage vendors is a very seamless move,” she adds.

Future enhancements are likely to continue the theme of pushing performance. “We get the most out of disks, we get the most out of networks and we get the most out of our own file system,” says Murphy, who expects those trends to continue to lead to higher performance.

“Panasas is unique,” she continues. “Other companies have focused on one or the other [large or small files]. We can do both very efficiently within one box.”

Prepare for pNFS

It has been a long time coming, but 2013 is the year that pNFS will gain adoption. Panasas has been working hard on developing an open-source pNFS Linux client, based on its proprietary DirectFlow protocol since 2007, and pNFS will be deployed in

standard Linux distributions this year.

pNFS solves many of the problems associated with file system performance encountered with legacy scale-up NAS systems. “pNFS enables breakthrough performance by allowing the compute clients to read and write data directly and in parallel to storage without the need for all I/O being funneled through a single congested filer head,” explains Murphy. “This allows storage systems architected for the pNFS standard to deliver significantly higher scalability and performance to keep up with the explosive growth in data most companies are experiencing.”

The commercial HPC market, in particular, will see substantial performance benefits. However, simply supporting the pNFS protocol will not guarantee the high performance that Panasas currently delivers with its ActiveStor storage solutions with DirectFlow. When it comes to pairing pNFS with the Panasas storage architecture and delivering the most performance, the object layout in pNFS has many advantages. Noer and his colleagues are convinced that Panasas will be ideally situated to continue to deliver perhaps the highest parallel performance in the industry. ■

www.panasas.com

Complete, massively scalable, open storage solutions for big data and HPC.

To the casual observer, the name Cray conjures up a picture of top-tier supercomputing centers. That tradition is still alive and well – the company still brands itself as “The Supercomputer Company” – with numerous installations on the semi-annual Top500 rankings, including the current #1 supercomputer at the Oak Ridge National Laboratory in Tennessee, appropriately named Titan.

But supercomputing is just one of three principle divisions at Cray, which is seeking to cement and expand its offerings in Big Data analytics and storage and data management.

Cray has actually been in the storage business for some time. To effectively unlock the performance capabilities of supercomputers and high performance computing (HPC) Linux clusters of all types requires scalable I/O. Customers requiring big I/O and fast data moved Cray into the parallel file systems and storage business. In 2003, Cray began investing heavily in parallel systems such as Lustre, which has matured to the leading open source parallel file system on the market. About two-thirds of the fastest systems in the top 100 use Lustre, and that number is growing.

In 2010, Cray and its storage partners founded OpenSFS, a consortium to advance Open Scalable File Systems—and specifically Lustre. Cray invests \$500,000 per-year to OpenSFS to improve Lustre. Cray also performs its own exhaustive testing, optimization, and enhancements to provide customers with a stable, hardware-optimized and certified version, called Cray Lustre File System (for component-based storage solutions) and also integrated with its Cray Linux Environment (CLE).

Today, Cray provides scalable storage solutions for Big Data and HPC. What’s less known is that Cray’s storage solutions fuel Linux clusters of all types. Cray’s best practices and architectures for Lustre (and other file systems) converge around three core product areas: storage systems, data management services, and data virtualization services.

In support of its systems business, Cray partners with DDN, NetApp, Xyratex, Mellanox, and others to provide complete, massively scalable, open storage solutions for Big Data and HPC. Cray may well be viewed as an incumbent in the storage and HPC

market, but the company has the feel of a start-up, says Jason Goodman, a senior manager in Cray’s storage team. “Projects are owned by people, not divisions and bureaucracy,” he says. Goodman was struck by the culture and Midwestern values of Cray’s employees when he joined the firm. “The people here are very down to earth, genuine, trustworthy,” he says, underlying the company’s tendency to under-promise and over-deliver.

Being known as trustworthy, storage vendor agnostic, and solutions-focused positions Cray nicely for Big Data and HPC markets needing more than just storage. Many customers need integrated solutions built on open systems that work as advertised. That differentiates Cray from other storage companies that lack systems expertise.

“The current trend in HPC is for a broader community of customers to be tackling more complex simulations with larger and faster data needs,” says Barry Bolding, Cray’s VP Storage and Data Management. “This positions Cray, with our expertise in scalable storage and compute, at the perfect place to meet the needs of this growing customer base, especially in the technical enterprise arena. Cray doesn’t have to go to Big Data -- Big Data is coming to Cray!”

Cray Pedigree

Cray was founded in the 1970s as Cray Research. It was sold to SGI, assets were acquired by Tera, and then reconstituted in the early 2000s. The company currently has close to 1,000 employees, with major R&D sites in Seattle, St Paul, and Austin. Many “Crayons” (Cray’s nickname for employees) left during the SGI days but came back when Cray assets were acquired by Tera, including the Cray name.

The Cray Storage and Data Management (SDM) division was created three years ago, growing organically from in-house expertise supplemented by storage experts from System Fabrics Works. The other two major pillars are Supercomputing (which embraces traditional supercomputing and commodity clusters running industry-standard Linux) and Yarc Data (which focuses on Big Data analytics, particularly graph analytics).

Historically Cray has been particularly strong in industries with massive data and I/O require-

ments: earth sciences (climate and weather prediction, geospatial imaging, and remote sensing), energy, government agencies and national labs. The company is growing in the life sciences arena, and gaining traction in sectors where “big I/O” is required and scaling beyond the bounds of traditional or scale-out NAS.

At its core, Cray is a holistic systems company. “We take pride in getting the customer set up in a way that works for them and being there for them,” says Goodman. There’s no such thing as dropping boxes of racks at the doorstep and walking away. We’re with the system into production.”

In addition to Cray’s own Sonexion scale-out Lustre storage system (see below), Cray offers component-based solutions built on the customer’s choice of storage. Cray provides optimized and validated solutions for DDN and NetApp using its Cray Lustre File System, also known as esFS (for external services file system). Cray also provides a suite of data connectivity and management products to provide data movement, cluster management, and interfaces to 3rd-party storage products such as HSM systems and archives.

These solutions offer customers tremendous flexibility, tailoring systems from terabytes (TB) to petabytes (PB). Cray optimizes the solution for any number of factors. For example, some customers require more performance with less capacity; others require more capacity and less performance. In other cases, customers need to scale performance and capacity in balanced increments as the storage system grows. For these customers, Cray offers an integrated approach -- Cray Sonexion.

Cray Sonexion System

Cray Sonexion embeds Lustre in an integrated, pre-assembled, -cabled, and -racked form-factor. The system is designed for larger environments where massive performance and capacity are required in a compact form factor. Cray Sonexion scales I/O from 5 GB/s to over 1 TB/s in a single file system.

The largest production installation of Cray Sonexion is at the University of Illinois for the NCSA Blue Waters Project.



Blue Waters provides over 26 PB across four file systems. 23 PB is in a single file system and used for a range of science and research applications, shared globally. At a large oil and gas company, Cray Sonexion attaches to a massive HPC IBM cluster, used for seismic modeling.

A big benefit to such petascale environments is footprint and complexity reduction. “What Cray has done is design and build a system that is huge,” said Michelle Butler, Technical Program Manager at Blue Waters. “I don’t know of any other system with the requirements for scalability that we are putting together for such huge applications. The performance will be above 1 TB/second... The machine can fill up this file system in a few hours.”

That is accomplished in about one-third of the footprint of conventional approaches to Lustre. It would take about 100 racks and over 5500 InfiniBand cables to do what Cray has done in 36 racks with about 500 cables.

Layer Cake

Cray’s solutions are gaining traction in a variety of industries with data-intensive and I/O-intensive workflows, where storage I/O and performance needs to scale with compute capability and capacity, and providing customers with choices in the process. Some prefer the integrated, appliance-like route, but others may opt for component-based solutions. But no matter what the choice, customers need systems that perform at scale and run predictably and reliably.

As parallel file systems continue maturing, more verticals and businesses will move Big Data workloads to parallel-class systems to utilize them for scalable performance, and supply more I/O. Big Data storage is not just about storing large volumes but dealing with I/O and memory issues, says Goodman. Some Big Data sets need Big I/O. “That’s what we’re in the business of—supplying I/O to enable massively concurrent or parallel processing and analysis,” says Goodman.

NFS is fantastic for simplicity and ubiquity but in cases where NAS is I/O or throughput bound, Cray can help. Cray provides NAS gateways and custom NFS solutions, but their core business

is delivering parallel storage solutions for Linux compute clusters of all types.

Not that many years ago, Goodman observes, many organizations made the leap from traditional NAS to scale-out architectures built on distributed file systems. “Now many Big Data workloads require even higher performance and capacity scalability. Thus, customers are turning to parallel-class systems to complement investments in NAS and prepare for massive data. It’s coming. Exascale is around the corner. Even parallel file systems will be put to their limits. Cray is building solutions for scaling I/O in lock-step with compute, using Cray DVS, SSD, and I/O forward techniques.”

Meanwhile, Cray is partnering closely with customers, deploying highly scalable storage solutions across industries. In support of its storage business, Cray invests in three core areas:

Leadership: Leading parallel file systems and storage architectures to unlock the potential of any HPC compute cluster and Big Data workload. These systems are not just under big supercomputers but attached to commodity Linux clusters as well. (Through its relationship with the National Laboratories, Cray collaborates on a number of exascale computing initiatives which will be broadly applicable to commercial Big Data in a very short time.)

Choice: Cray does what’s right for customers and builds on open systems. That means providing customers with choice of storage hardware, flexibility, and foundation in open systems.

Results: Cray ultimately measures its effectiveness through results. Says Goodman: “It’s about keeping up with science, research, exploration--and compute. In the Cray storage world, we measure sustained performance, but at the end of the day we do everything we can to help customers get results faster.” ■

Further Reading: Cray White Paper: “Enabling Scientific Breakthroughs at the Petascale.” http://www.sonexion.com/whitepapers/Cray_Blue_Waters.pdf

www.cray.com

Solving the tough problems at the intersection of high performance and high capacity

DataDirect Networks (DDN) is a leading provider of high-performance, high-capacity storage environments, but according to Jeff Denworth, VP of Marketing, DDN is going way beyond that. With a unique and exacting focus on the requirements of today's massive unstructured data generators, DDN innovations across block, file, object and cloud solutions have driven the company's success in thousands of world's most demanding data environments including top Web 2.0 companies, large content providers and over two thirds of the world's fastest computers.

nization," says Denworth. Put another way, it is the intersection of high performance with scalability, reliability and manageability that has been the company's primary focus.

DDN's early growth was fueled by an engineering focus on handling large data rates as well as volumes, for high-performance computing. "That's been part of our DNA from the get-go," comments Denworth.

An early customer was NASA, which selected DDN because it was the only vendor capable of ingesting the very high rates of data being transmitted. DDN subsequently added data process-

“Computing is being transformed, new companies are emerging. Many organizations that have Big Data don't have the ability to process Big Data. ”

In many ways, DDN has been mastering Big Data challenges before the term 'Big Data' was even invented. By supporting the requirements of the world's largest file storage systems and demanding applications for over 10 years, DDN has developed both domain expertise and leadership in today's new data explosion.

DDN is the third company built by the two co-founders – Alex Bouzari and Paul Bloch. "They came together with a real vision for what they wanted to accomplish," says Denworth. United in the belief that data storage isn't just a commodity, DDN's co-founders aimed to commercialize a true high-end massively scalable data storage platform.

Time to Insight

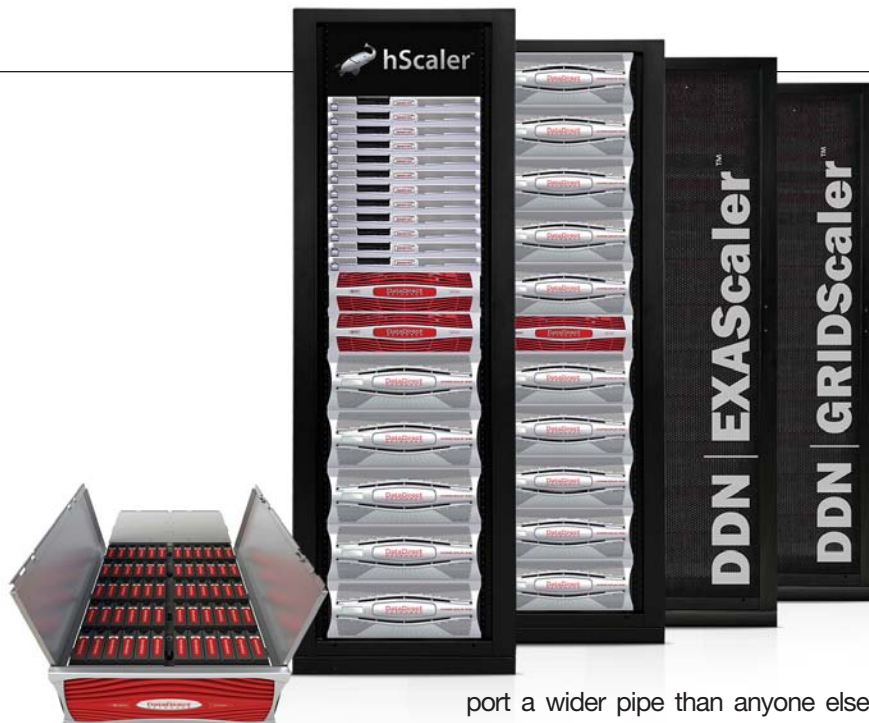
Although headquartered in Los Angeles, in many respects DDN has the DNA of a Silicon Valley start-up. For DDN, "Big Data" is not just about the volume of data but what users can do with it. "It's the time to insight for an orga-

ing capability, becoming the first storage vendor to develop an in-storage processing engine.

What is the secret of DDN's success? The company's competitive strengths owe a lot to engineering, including great flexibility at the system level.

For one thing, DDN is truly agnostic when it comes to file systems. "Many storage vendors restrict the choice of file system for their customers," Denworth points out. If the hardware or vendor support is tied exclusively to one file system the customer is locked-in and risks losing their investment in the face of changing requirements.

For example, DDN actively supports and has invested more than \$10 million in R&D for the popular Lustre file system. But DDN also supports many leading commercial file systems, including GPFS. DDN is also agnostic when it comes to interconnects, and supports multiple formats, including Fibre Channel, Infiniband, and Ethernet.



And then there is DDN's pay-as-you-grow approach and "green IT" philosophy, which provides scale without overtaxing data center resources. The density of the storage systems affords straightforward scaling from <100TB into the multi-petabyte range. "With industry-leading density, you get 3.4PB of storage per rack giving the architecture inherent efficiency on space, power and cooling".

Another reason behind the company's growth is that every system is built, assembled, and tested by DDN as a complete system. "We do extensive system-level integration and testing, ensuring quality at scale," says Denworth. That in turn makes for a good customer experience and ease of deployment. This is especially important for parallel file systems where vendor integration experience can save users significant time in deployment and tuning.

Market Reach

DDN has roughly doubled its size in the past two years, seeing particularly strong growth in the HPC market, its traditional core market. According to Denworth, the company currently supports about over two thirds of the Top100 supercomputers and delivers more bandwidth than all other Top500 storage providers combined.

Other strong growth areas include Web 2.0 applications – a testament to DDN's ability to not only ingest data but also distribute data out of storage. "We can sup-

port a wider pipe than anyone else, so that matters for a content provider who has to transcode assets regularly into over ten different formats – sometimes in close to real time," says Denworth, referring to content providers' need to distribute video data to many different devices each requiring different encoding methods. "Content providers have many storage challenges, not only the demanding performance of transcoding, but also the ability to reliably distribute assets locally, nationally or globally at high speed and quality," says Denworth.

Another important growth area for DDN is government, particularly digital surveillance and security, boosted by higher-resolution video technologies for monitoring around the world. And DDN is reporting strong growth in financial services and the manufacturing sector, as new modeling and simulation technologies push high growth of HPC element in those industries.

Indeed, few competitors have the experience of dealing with tens of Petabytes in a customer environment, says Denworth. "We lean on our credibility – for true big data applications, we've been the leader for some time," he says.

DDN has a reputation for innovation in the use of storage in a workflow: the company was the first to deploy in-storage compute capabilities, and was also the first to create a cache-centric architecture to manipulate large data sets at high speed.

DDN now carries their innovation and big data leadership into the area

of analytics where DDN solutions have delivered the world record performance for SAS GRID, Vertica and Kx, and outperforms commodity hardware by 2 to 7x in Hadoop environments.

The performance work in analytics led DDN to develop one of their most exciting new products; the hScaler appliance. The product combines all of the key elements required to manage and run a Hadoop environment in an appliance form with the high-throughput storage that DDN is famous for to accelerate map-reduce computation.

"hScaler has strong appeal across both traditional HPC and Enterprise HPC operations," says Denworth. "The strongest demand today comes from Genomics, Pharma, Financial Services and Government, but there is rising interest also in our other core markets." DDN's forte, in life sciences and other domains, is impacting areas where traditional IT environments don't accelerate customers' efforts.

Four Pillars

In fact, when it comes to Big Data, DDN pictures four major pillars to its storage architecture: ingest, processing, distribution and storage. DDN has innovated in all four areas, all with a common purpose of accelerating customers' work and research.

"You can't just think in the old terms of block storage," says Denworth. We're unique in being able to bring together in single workflows SAN, NAS and object-based storage platforms from DDN. You can't approach Big Data with just one arrow in your quiver. The same way we're file-system and interconnect agnostic, we're also format agnostic in terms of process."

Despite this period of rapid technological innovation, "People are approaching Big Data with old technologies," says Denworth. "Computing is being transformed, new companies are emerging. Many organizations that have Big Data don't have the ability to process Big Data."

That's a core issue – one that DDN's massively scalable storage is tackling with considerable success. ■

www.ddn.com